GHAJAR
EXHIBIT 45

Page 1

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

_____

RICHARD KADREY, et al.,   )
)
          Individual and   )
          Representative   )
          Plaintiffs,   )
)
v.   )   Case No. 3:23-cv-03417-VC
)
META PLATFORMS, INC.,   )
)
          Defendant.   )
_____)

    ** HIGHLY CONFIDENTIAL - ATTORNEYS' EYES ONLY **

        Videotaped Deposition of EUGENE NHO

            San Francisco, California

            Friday, December 6, 2024

        Reported Stenographically by

    Michael P. Hensley, RDR, CSR No. 14114

_____

            DIGITAL EVIDENCE GROUP

          1730 M. Street, NW, Suite 812

            Washington, D.C. 20036

              (202) 232-0646

Page 39

 1   THE WITNESS: I think I answered the

 2   question earlier, which was you do think about it.

 3   It is very hard to imagine whether it would lose

 4   money or not because you don't know who would adopt

 5   it. It was an entirely new attempt.

 6   We have never done anything like that; so

 7   there hasn't been -- I don't recall that as a very

 8   prominent, salient factor in price determination.

 9   BY ATTORNEY STOLER:

10   Q.   All right. And you said you moved to a

11   position in GenAI in May; correct?

12   A.   That is correct.

13   Q.   All right. What is your position in the

14   GenAI group?

15   A.   Product manager.

16   Q.   So you've been a product manager your

17   entire time at Meta; correct?

18   A.   That is correct.

19   Q.   All right. Have you received any

20   promotion during your time at Meta?

21   A.   I have.

22   Q.   All right. But that promotion did not

Page 40

1   come with a change in title; correct?

2   A.   It did.   I -- I'm a director of product

3   management, but product manager is the -- PM is the

4   function that I'm in; so I don't necessarily change

5   what I call myself. I'm a product manager. I could

6   be a lead product manager, staff product manager,

7   director of product management, but I'm product

8   manager.

9   Q.   All right. Well, if you could describe to

10   me -- let's start here.

11   What -- what's your current title at Meta?

12   A.   Official title?

13   Q.   Yes.

14   A.   Director of product management.

15   Q.   All right. And how long have you been a

16   director of product management?

17   A.   Roughly nine months.

18   Q.   So were you promoted to a director of

19   product management while you were in the

20   monetization group?

21   A.   That is correct.

22   Q.   All right. And then shortly after that

12/6/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Eugene Nho
                      Highly Confidential - Attorneys' Eyes Only

Page 43

 1   A.   FAIR is a research organization that is a

 2   separate sister org to the entire GenAI org.

 3   Q.   And what does FAIR work on?

 4   A.   They work on fundamental research in

 5   various domains across AI that may or may not have

 6   to do with generative AI technologies.

 7   Q.   All right. And so in your time in the

 8   GenAI group, have you always been in the

 9   foundational team?

10   A.   That is correct.

11   Q.   All right. And I think you previously

12   testified that that foundational team is responsible

13   for developing the Llama models; is that correct?

14   A.   That is correct.

15   Q.   Do you have any role in developing the

16   Llama models?

17   A.   I do.   I work on data and currently also

18   reasoning.

19   Q.   What does it mean to work on reasoning?

20   A.   Teach the model to be really good at

21   reasoning so you can solve very complex math

22   problems or physics problems, et cetera, et cetera.

Page 44

1   Q.   All right.   What are your responsibilities

2   in your current role in the foundational team?

3   A.   I set the long-term vision for data for

4   Llama; so various ways and places where we need to

5   get the best data in the world to train the Llama

6   models, and for this other responsibility which is

7   relatively new.

8   For reasoning, I think about what ultimate

9   use cases we want to be good at with reasoning and

10   how we should consequence our research activity so

11   that we can get there.

12   Q.   Okay.   You mentioned that you set the

13   long-term vision for data for Llama; right?

14   A.   That is correct.

15   Q.   All right.   And that involves, in your

16   words, getting the best data in the world to train

17   the Llama models; correct?

18   A.   That is correct.

19   Q.   All right.   And when you say "get the best

20   data," what -- what do you mean by that?

21   A.   There are a number of different levers, or

22   channels, to get data.   One obvious place is our own

Page 45

1   first party data, because we have a lot of our

2   on-platform user data.   There's crawling.   There is

3   synthetic data, so we use the model to generate

4   data.   There is licensing.

5   And there are a few others that are things

6   like how do you train the model to be equally good

7   but not -- without using as much data; so how do you

8   do more with less?   Annotation, et cetera.

9   So there's a range of levers that we need

10   to use to acquire the best data for Llama models.

11   Q.   All right. One method of data acquisition

12   you just mentioned was crawling; right?

13   A.   That is correct.

14   Q.   All right. Can you tell me what crawling

15   is.

16   A.   In layman's words, it's you go and capture

17   the web. It's -- you basically automatically --

18   just like a human would read a web page, imagine a

19   crawler, or a machine, that goes in, reads all the

20   webs pages, and downloads that.

21   The same thing that Google does in order

22   to power their search engines. The same thing that

Page 85

1   Q.   Okay.   And you said you try to develop

2   intuitions about what is and isn't quality data;

3   right?

4   A.   I did, yeah.

5   Q.   And are there any metrics that inform

6   those intuitions about what is and isn't

7   high-quality data?

8   ATTORNEY BRIGHAM: Objection.   Form.

9   THE WITNESS: That's the ultimate

10   benchmark, the model performance benchmark.   There

11   are a set of industry benchmarks that measure

12   whether the model is great at.   Does it know these

13   kind of knowledge questions well?   Does it know how

14   to solve math problems?   Does it know how to do

15   graduate level across the domain tasks?   Is it good

16   at multilingual tasks, et cetera?

17   So there are all these metrics, or

18   benchmarks, that the industry has that many of the

19   LLMs run on those benchmarks to see how they

20   perform.   That's what -- what eventually we look at.

21   But you can't just look -- it's -- again,

22   that's why that ablation step, testing step, is

Page 86

1   important; because until we do it, it's hard for us

2   to predict ultimately how good the model would be.

3   BY ATTORNEY STOLER:

4   Q.   So in a situation where your -- you do

5   have possession of the data such that you could run

6   an ablation study, you would use industry benchmarks

7   to form an intuition of whether or not that data is

8   of high quality?

9   ATTORNEY BRIGHAM: Objection. Form.

10   THE WITNESS: Actually, that is incorrect.

11   The ablation itself, the test is -- the outcome of

12   the test that the industry -- industry benchmark

13   score, you have specific data, you have that

14   version -- you have a -- you have a model that is

15   trained on with this data and you have a model that

16   is trained on without this data, and you look at how

17   it performs in those set of industry benchmarks.

18   You look at benchmark and that's how you determine

19   whether this data was good or not.

20   BY ATTORNEY STOLER:

21   Q.   But in a situation where you don't have

22   the data available to you --

Page 117

1   version of it where we would talk about across

2   different modalities, what should we be doing.

3   So does that answer your question?

4   Q.   No.

5   So I'm asking just if you know one way or

6   the other of a document referred to as a "text

7   pretraining data road map."

8   A.   I know it exists. I think I've seen it.

9   I can't tell you what's in it right now.

10   Q.   Okay. That does answer my question.

11 Thank you.

12   All right. If you'd turn now to the

13 second page of Exhibit 2. And take a look -- these

14 headers are numbered here, and if you'd look with me

15 to number 6 which starts "How to address the

16 research risk."

17   Do you see that?

18   A.   I do.

19   Q.   Do you know what research risk is being

20 referred to in this header?

21   A.   One second. Let me just take a look.

22   I do.

Page 118

1   Q.   And can you tell me what that is?

2   A.   This is talking about how from -- going

3   from Llama 3 to Llama 4, we need to have far larger

4   data size, and then you need -- when you have a

5   larger data size, the specific proportions of

6   data -- the -- make sure that the data proportions

7   stay in an optimal way is really hard.

8   I guess another way -- easier way to think

9   about it is we're -- we're constructing a learning

10   curriculum for the model, and when the learning

11   curriculum has to become multiple times larger, then

12   making sure that the right kinds of the topics get

13   covered in the right proportion gets harder because

14   certain topics are really hard to source, and they

15   just don't exist that many -- not much exists in the

16   world.

17   Does that make sense?

18   Q.   Yes. Thank you.

19   All right. You can put Exhibit 2 aside

20   for the moment. We'll return to it.

21   A.   Mm-hmm.

22   ATTORNEY STOLER: All right. I'm going to

Page 148

```
 1              ATTORNEY STOLER:  We can -- well --

 2    BY ATTORNEY STOLER:

 3         Q.   All right.  And then it says, "or when we

 4    identify datasets with particularly high expected

 5    value."

 6              You see that?

 7         A.   I see that.

 8         Q.   Okay.  Do you know of any situation where

 9    you've identified datasets with particularly high

10    expected value?

11         A.   Yes.  There are actually a number of

12    circumstances when we would get introductions, or we

13    would somehow run into a potential partner who has

14    data that's valuable.  And then we decide to have --

15    see if there's a potential licensing path.

16              It's -- a lot of times, there isn't.

17    Like, they don't even know how to think about

18    pricing.  They don't even know a lot of times

19    whether they have the rights to license or not.  But

20    we would -- an example would be, you know, we got

21    introduced to a company that runs call centers or

22    builds a technology that powers call centers, and
```

Page 149

1   they might have really interesting data to train our

2   speech models.

3   We got introduced a agency that helps

4   Indian students pick their college majors and

5   prepare them for college entrance exams. They had

6   some interesting test prep data, and this was

7   through, like, a personal connection of the

8   employee. And we ran into them. It seems like they

9   have high-quality data; so we pursued that.

10   So this is basically saying like

11   opportunistically, we'll also -- as we find them,

12   we'll pursue them.

13   ATTORNEY STOLER: Got it. And --

14   Why don't we go off the record.

15   THE VIDEOGRAPHER: Okay. We're off the

16   record at 1:56.

17   (A break was taken.)

18   THE VIDEOGRAPHER: We are now on the

19   record at 3:09.

20   BY ATTORNEY STOLER:

21   Q.   All right, Mr. Nho. Picking up where we

22   left off, we were discussing the -- what is, I

Page 248

 1   CERTIFICATE OF SHORTHAND REPORTER

 2

 3   I, Michael P. Hensley, Registered Diplomate

 4   Reporter for the State of California, CSR No. 14114,

 5   the officer before whom the foregoing deposition was

 6   taken, do hereby certify that the foregoing

 7   transcript is a true and correct record of the

 8   testimony given; that said testimony was taken by me

 9   stenographically and thereafter reduced to

10   typewriting under my direction; that reading and

11   signing was not requested; and that I am neither

12   counsel for, related to, nor employed by any of the

13   parties to this case and have no interest, financial

14   or otherwise, in its outcome.

15

16

17

18

19

20

21   Michael P. Hensley, CSR, RDR

22